Modeling Latency and Shape Changes in Trial Based Neuroimaging Data

Mørup, Morten; Hansen, Lars Kai; Madsen, Kristoffer Hougaard

Published in:
2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)

Link to article, DOI:
10.1109/ACSSC.2011.6190037

Publication date:
2011

Modeling Latency and Shape Changes in Trial Based Neuroimaging Data

Morten Mørup and Lars Kai Hansen
Section for Cogntive Systems,
DTU Informatics, Technical University of Denmark
{mm,lkh}@imm.dtu.dk

Kristoffer Hougaard Madsen
Danish Research Centre for Magnetic Resonance
Copenhagen University Hospital Hvidovre
khm@drcmr.dk

***Abstract*—To overcome poor signal-to-noise ratios in neuroimaging, data sets are often acquired over repeated trials that form a three-way array of *space*×*time*×*trials*. As neuroimaging data contain multiple inter-mixed signal components blind signal separation and decomposition methods are frequently invoked for exploratory analysis and as a preprocessing step for signal detection. Most previous component analyses have avoided working directly with the tri-linear structure, but resorted to bi-linear models such as ICA, PCA, and NMF. Multi-linear decomposition can exploit consistency over trials and contrary to bi-linear decomposition render unique representations without additional constraints. However, they can degenerate if data does not comply with the given multi-linear structure, e.g., due to time-delays. Here we extend multi-linear decomposition to account for general temporal modeling within a convolutional representation. We demonstrate how this alleviates degeneracy and helps to extract physiologically plausible components. The resulting convolutive multi-linear decomposition can model realistic trial variability as demonstrated in EEG and fMRI data.**

I. Introduction

Analysis of neuroimaging data sets, say electroencephalography (EEG), magnetoencephalography (MEG) or functional magnetic resonance imaging (fMRI), is hampered by noise, confounds, and the presence of multiple mixed signal components of interest. To overcome poor signal to noise (SNR) data is typically averaged over repeated trials. In this contribution we propose a flexible averaging approach based on a convolutive multi-linear model. We show that the new method provides for improved representations of fMRI and EEG data relative to two competing multi-linear decomposition methods based on instantaneous mixing and mixing with single delays.

EEG and fMRI data may be represented by a *space* × *time* matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$ with elements $x_{i,j}$. Bi-linear component analyses are routinely applied to neuroimaging data for exploratory investigations or as a pre-processing step prior to signal detection, see for instance [8], [19], [20]. The bi-linear model reads

$$x_{i,j} \approx r_{i,j} = \sum_{d=1}^{D} a_{i,d} b_{j,d},$$

where the data is represented as a sum of components with time profiles $\mathbf{b}_1, \ldots, \mathbf{b}_D$ and corresponding spatial topographies $\mathbf{a}_1, \ldots, \mathbf{a}_D$. Since such factor analytic representations are ambiguous, additional constraints may be imposed. For singular value decomposition (SVD) and principal component analysis (PCA) profiles are assumed orthogonal as eigenvectors of the covariance matrix, while for independent component analysis (ICA) an independence assumption is imposed for one of the two modes.

When data are recorded over K repeated trials, we obtain a *space*×*time*×*trials* hypermatrix $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ also commonly denoted a multi-way array or tensor. Bi-linear methods first convert the tri-linear form to a matrix, either by 'matricizing' ($\mathcal{X}^{I \times J \times K} \rightarrow \mathbf{X}^{I \times JK}$) or by averaging trials, thus ignoring the inherent structure of the repeated trials. By working directly on the tri-linear structure we can both allow for a trial-dependent weight (contrary to simple averages) as well as impose consistency in the time-profiles (contrary to bi-linear decompositions of the matricized array) through the CANDECOMP/PARAFAC (CP) model [5], [11] $x_{i,j,k} \approx r_{i,j,k} = \sum_{d=1}^{D} a_{i,d} b_{j,d} c_{k,d}$. Here $\mathbf{c}_d$ represents the strength in which the profile time series $\mathbf{b}_d$ with spatial topography $\mathbf{a}_d$ is expressed. Kruskal proved in [17] that the CP model is unique under mild conditions that in general are satisfied in the presence of noise. Consequently, modeling repeated trials by CP in theory resolves the ambiguities encountered when modeling the data by (bi-linear) factor analysis. The application of CP to EEG was first suggested in [11] and was later reinvented in [21] under the name *topographic component analysis*. In [1] it was further demonstrated how the CP model is useful in the analysis of fMRI.

While the tri-linear model can accommodate trial-to-trial variability of amplitude it can not model the inevitable inter-trial delay and shape variations of typical neuroimaging experiments, see e.g., [25]. To accommodate such trial variability we propose the to use the convolutive CP model[22], which we denote convCP,

$$x_{i,j,k} \approx r_{i,j,k} = \sum_{d=1,\tau=1}^{D,T} a_{i,d} b_{j-\tau,d} c_{k,d,\tau}. \qquad (1)$$

Here, the time profile $\mathbf{b}_d$ is present with delay τ in trial k with strength $c_{k,d,\tau}$. The proposed model is a generalization of the model analyzed in [15], [16], [24], which we here will denote shiftCP. The shiftCP model is constrained such that each time profile only has one specific delay value in each trial. Thus, contrary to shiftCP the convCP model allows for an arbitrary number of possible component delays at each trial within the length T of the convolutive filter. The CP, shiftCP and convCP models are illustrated in figure 1.

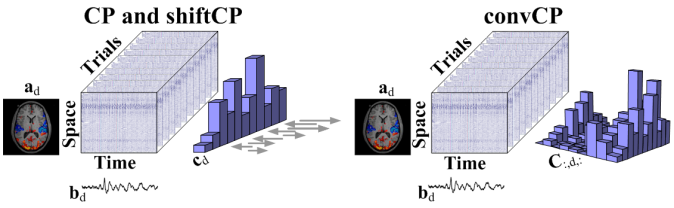

Fig. 1. Illustration of the CP and shiftCP model (left) as well as convCP model (right). The CP model decompose the *space*×*time*×*trials* hyper-matrix into the component profiles $\mathbf{a}_d$, $\mathbf{b}_d$,$\mathbf{c}_d$ pertaining to the *space*, *time* and *trials* modality respectively. The shiftCP model further allow for a specific delay for each component in each trial indicated by the gray arrows. The convCP model also decompose the data into component profiles pertaining to each modality, however, an arbitrary number of delays are possible within the complete convolutive filter $\mathbf{C}_{:,\mathbf{d},:}$. Furthermore, filtering the time series signal through the convolutive filter also allow for shape variability.

Data generated from the convCP model are no longer multi-linear rendering a CP decomposition invalid. When data violate multi-linearity, 'CP-degenerate' solutions are known to occur. Roughly speaking, this refers to solutions in which some component loadings are highly correlated in all modes and the elements of these components can become arbitrarily large. CP-degeneracy makes the estimation unstable, causes slow converge, and renders the results difficult to interpret – largely because the model is plagued by strong between-component cancelations [12], [7]. To avoid CP-degeneracy, constraints in the form of orthogonality [12], [9] or independence [2] have been imposed. As degeneracy is mainly an indication of model inadequacy, we avoid restricting the CP model, and rather seek to expand the model to incorporate shape and delay changes through the above convCP representation.

We note that convolutive bi-linear decomposition has been extensively studied in the context of ICA and NMF, see e.g., [27], [26]. Bi-linear convolutive decompositions invariant to shift have also been proposed within the framework of sparse coding, by imposing sparsity constraints on the component time series [18], [4]. Furthermore, bi-linear analysis with delayed mixing have been treated in numerous papers, see for instance [3], [28], [29]. However, convolutive multi-linear modeling of neuroimaging data has to our knowledge not been treated previously.

II. The convCP model

Let $\mathbf{A}^{I\times D}$, $\mathbf{B}^{J\times D}$ and $\mathcal{C}^{K\times D\times T}$ be (hyper-)matrices holding the profiles defined in the convCP model in equation (1). In the following $\mathcal{C}$, $\mathbf{B}$ and $\tilde{\mathcal{C}}$, $\tilde{\mathbf{B}}$ will denote the same (hyper-)matrix in the time and frequency domain respectively, i.e. $\tilde{\mathcal{U}} = \mathfrak{F}(\mathcal{U})$, $\mathcal{U} = \mathfrak{F}^{-1}(\tilde{\mathcal{U}})$ using the discrete fourier transform (DFT) and inverse fourier transform along the modality indexed by $j \in \{1, 2, ..., J\}$ and $\tau \in \{1, 2, ..., T\}$ respectively. Notice, when $T < J$, $\mathcal{C}$ is zero padded along the third mode to have size J. $\mathcal{C}_{:,:,1:T}$ will denote the entries $\{1, 2, ..., T\}$ of the third modality, while $\mathcal{C}_{:,d,:}$ will denote the d^{th} slab off the second mode – in particular we denote $\mathbf{C}^{\tau} = \mathcal{C}_{:,:,\tau}$. We further define $\mathbf{B}(\tau)$ such that $b(\tau)_{j,d} = b_{j-\tau,d}$. Finally, $*$ will denote the complex conjugate, $\mathbf{D}_{\mathbf{A}_{i,:}}$ a diagonal matrix containing the i^{th} row of $\mathbf{A}$ along the diagonal, $\theta(.)$ the unit step function and $\mathbf{A}^t$ the variable $\mathbf{A}$ at the t^{th} iteration.

A. Uniqueness of ConvCP representations

Kruskal's rigorous general proof of CP-uniqueness is involved [17]. A simpler proof of uniqueness can be obtained if $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ all have full rank [13], [21]. For the i^{th} slab the CP model reads

$$\mathbf{X}_{i,:,:} \approx \mathbf{B}\mathbf{D}_{\mathbf{A}_{i,:}}\mathbf{C}^{\mathsf{T}} = \mathbf{B}\mathbf{P}[\mathbf{P}^{-1}\mathbf{D}_{\mathbf{A}_{i,:}}\mathbf{Q}]\mathbf{Q}^{-1}\mathbf{C}^{\mathsf{T}} = \mathbf{B}'\mathbf{D}_{\mathbf{A}'_{i,:}}\mathbf{C}'^{\mathsf{T}}.$$

Thus, if two solutions $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{A}', \mathbf{B}', \mathbf{C}'$ exist there must be a mapping from one solution to the other given by $\mathbf{P}$ and $\mathbf{Q}$. However, for this mapping the term $\mathbf{P}^{-1}\mathbf{D}_{\mathbf{A}_{i,:}}\mathbf{Q}$ has to be diagonal for all i. When $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ have full rank this restricts $\mathbf{P}$ and $\mathbf{Q}$ to be simple scale and permutation matrices [13], [21]. For the convCP model we have

$$\begin{aligned}\mathbf{X}_{i,:,:} &\approx \sum_{\tau} \mathbf{B}(\tau)\mathbf{D}_{\mathbf{A}_{i,:}}\mathbf{C}^{\tau^{\mathsf{T}}} = \sum_{\tau} \mathbf{B}(\tau)\mathbf{P}[\mathbf{P}^{-1}\mathbf{D}_{\mathbf{A}_{i,:}}\mathbf{Q}]\mathbf{Q}^{-1}\mathbf{C}^{\tau^{\mathsf{T}}} \\ &= \sum_{\tau} \mathbf{B}'(\tau)\mathbf{D}_{\mathbf{A}'_{i,:}}\mathbf{C}'^{\tau^{\mathsf{T}}}\end{aligned}$$

Although, the convCP model extends the CP model $\mathbf{P}^{-1}\mathbf{D}_{\mathbf{A}_{i,:}}\mathbf{Q}$ must remain diagonal for all values of i. An ambiguity is however introduced in the convCP model between the component time series $\mathbf{b}_d$ and convolutive filter $\mathcal{C}_{:,d,:}$ such that the temporal profiles to some extent can be coded arbitrarily in either the filter coefficients or in the component time series. Furthermore, the convCP model is equivalent to bilinear decomposition when the length of the convolutive filter is the same as the number of sampled time-points $T = J$. To overcome these ambiguities we impose – in line with e.g., [18], [4] - a sparse prior distribution on the filter coefficients (rather than component time series) such that the time signature is coded in the component time series while delays and shape variations are coded in the (sparse) filter.

B. Model estimation

We formulate the model in a probabilistic framework, which in future work will allow us to use standard Bayesian tools to optimize hyper-parameters and perform model selection (see also [23]), here we focus on algorithmic issues for MAP estimation of $\mathbf{A}, \mathbf{B}, \mathcal{C}$. We assume normal i.i.d. noise, $P(\varepsilon_{i,j,k}) \sim N(0, \sigma^2)$, thus the likelihood function is $P(\mathcal{X}|\mathbf{A}, \mathbf{B}, \mathcal{C}, \sigma^2) = \prod_{i,j,k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{e^2_{i,j,k}}{2\sigma^2}]$, where we use the shorthand e for the residual. To enforce consistency over trials of the components we assign the sparse positive i.i.d. exponential prior for the filter coefficients $P(c_{k,d,\tau}|\lambda) = \lambda \exp[-\lambda c_{k,d,\tau}]\theta(c_{k,d,\tau})$. To alleviate the scale ambiguity inherent in the tri-linear model we finally assign uniform priors over the unit hypersphere for the columns of $\mathbf{A}$ and $\mathbf{B}$: $P(\mathbf{A}) \propto \prod_d \delta(\|\mathbf{a}_d\|_F - 1)$, $P(\mathbf{B}) \propto \prod_d \delta(\|\mathbf{b}_d\|_F - 1)$. Using Bayes theorem the joint posterior for $\mathbf{A}$, $\mathbf{B}$ and $\mathcal{C}$ can be written as $P(\mathbf{A}, \mathbf{B}, \mathcal{C}|\mathcal{X}, \sigma^2, \lambda) \propto P(\mathcal{X}|\mathbf{A}, \mathbf{B}, \mathcal{C}, \sigma^2)P(\mathbf{A})P(\mathbf{B})P(\mathcal{C}|\lambda)$. Ignoring constants and subjecting $\|\mathbf{a}_d\|_F = \|\mathbf{b}_d\|_F = 1$, the negative log-posterior is given by $-\log P(\mathbf{A}, \mathbf{B}, \mathcal{C}|\mathcal{X}, \sigma^2, \lambda) = \frac{1}{2\sigma^2}\sum_{i,j,k} \|x_{i,j,k} -$

Algorithm 1 Estimation of the parameters of the Convolutive CP

1: Initialize $\mathbf{A}^0$, $\mathbf{B}^0$, $\mathcal{C}^0$ by random such that $\|\mathbf{a}_d^0\|_F = \|\mathbf{b}_d^0\|_F = 1$ and $c_{k,d,\tau}^0 \geq 0$
2: **repeat**
3: **A-update:** $g_{i,d} = \sum_{k,j,\tau} b_{j-\tau,d}^t c_{k,d,\tau}^t (x_{i,j,k} - \sum_{d',\tau'} a_{i,d'}^t b_{j-\tau',d'}^t c_{k,d',\tau'}^t)$, $g_{i,d} = g_{i,d} - a_{i,d}^t \sum_i g_{i,d} a_{i,d}^t$
$\mathbf{A}^{t+1} = \mathbf{A}^t - \mu_A \mathbf{G}$, $\mathbf{a}_d^{t+1} = \frac{\mathbf{a}_d^{t+1}}{\|\mathbf{a}_d^{t+1}\|_F}$, estimate μ_A by line-search such that $P(\mathbf{A}^{t+1}, \mathbf{B}^t, \mathcal{C}^t | \mathcal{X}, \lambda') > P(\mathbf{A}^t, \mathbf{B}^t, \mathcal{C}^t | \mathcal{X}, \lambda')$
4: **B-update:** $\widetilde{g}_{f,d} = \sum_{k,i} a_{i,d}^{t+1} \widetilde{c}_{k,d,f}^{*t} (\widetilde{x}_{i,f,k} - \sum_{d'} a_{i,d'}^{t+1} \widetilde{b}_{f,d'}^t \widetilde{c}_{k,d',f}^t)$, $\mathbf{G} = \mathfrak{F}^{-1}(\widetilde{\mathbf{G}})$, $g_{j,d} = g_{j,d} - b_{j,d}^t \sum_j g_{j,d} b_{j,d}^t$
$\mathbf{B}^{t+1} = \mathbf{B}^t - \mu_B \mathbf{G}$, $\mathbf{b}_d^{t+1} = \frac{\mathbf{b}_d^{t+1}}{\|\mathbf{b}_d^{t+1}\|_F}$ estimate μ_B by line-search such that $P(\mathbf{A}^{t+1}, \mathbf{B}^{t+1}, \mathcal{C}^t | \mathcal{X}, \lambda') > P(\mathbf{A}^{t+1}, \mathbf{B}^t, \mathcal{C}^t | \mathcal{X}, \lambda')$
5: $\mathcal{C}$**-update:** $\widetilde{g}_{k,d,f} = \sum_i a_{i,d}^{t+1} \widetilde{b}_{f,d}^{*t+1} (\widetilde{x}_{i,f,k} - \sum_{d'} a_{i,d'}^{t+1} \widetilde{b}_{f,d'}^{t+1} \widetilde{c}_{i,d',f}^t)$, $\mathcal{G} = \mathfrak{F}^{-1}(\widetilde{\mathcal{G}}) + \lambda'$
$\mathcal{C}^{t+1} = \mathcal{C}^t - \mu_C \mathcal{G}_{:,:,1:T}$, $\mathcal{C}_{k,d,\tau}^{t+1} = \mathcal{C}_{k,d,\tau}^{t+1} \theta(\mathcal{C}_{k,d,\tau}^{t+1})$, estimate μ_C by line-search such that $P(\mathbf{A}^{t+1}, \mathbf{B}^{t+1}, \mathcal{C}^{t+1} | \mathcal{X}, \lambda') > P(\mathbf{A}^{t+1}, \mathbf{B}^{t+1}, \mathcal{C}^t | \mathcal{X}, \lambda')$
6: **until** convergence

$r_{i,j,k}\|_F^2 + \lambda \sum_{k,d,\tau} |c_{k,d,\tau}| + const$. In the following we define $\lambda' = \lambda\sigma^2$. Using Parseval's identity and making the standard approximation invoking a circular convolution the DK convolutions can be written as a complex multiplication reducing the computational complexity of each convolution from $\mathcal{O}(JT)$ to $\mathcal{O}(J \log J)$, i.e. $\frac{1}{2} \sum_{i,j,k} \|x_{i,j,k} - \sum_{d,\tau} a_{i,d} b_{j-\tau,d} c_{k,d,\tau}\|_F^2 = \frac{1}{2J} \sum_{i,f,k} \|\widetilde{x}_{i,f,k} - \sum_d a_{i,d} \widetilde{b}_{f,d} \widetilde{c}_{k,d,f}\|_F^2$. The algorithm based on maximum a posteriori estimation for convCP is outlined in Algorithm 1. The $\mathbf{B}$ and $\mathcal{C}$ updates were sped up by pre-computing $\sum_i \widetilde{x}_{i,f,k} a_{i,d}$ and $\sum_i a_{i,d} a_{i,d'}$. To enforce the normalization of $\mathbf{A}$ and $\mathbf{B}$ we re-normalized the variables at each iteration and calculated the gradient such that it is invariant to the normalization – this follows by the steps $g_{i,d} = g_{i,d} - a_{i,d} \sum_i g_{i,d} a_{i,d}$ and $g_{j,d} = g_{j,d} - b_{j,d} \sum_j g_{j,d} b_{j,d}$ in the $\mathbf{A}$ and $\mathbf{B}$ update. To capture components strictly time locked to the event it is possible to let the filter coefficients for $\tau \in \{2, 3, ..., T\}$ of one or several of the components be zero. To achieve this the corresponding gradient for these filter coefficients are also set to zero. Furthermore, as no pruning is desired for these components we here assume a uniform non-negative prior ($\lambda \to 0$). In the following we will denote the uniform non-negative prior by $\lambda = 0$. Notice that in order to take advantage of the Fourier transform we require the source time course in each trial to be periodic, if this is not the case approximate periodicity can be achieved by use of a window function.

III. Experimental Results and Discussion

We first analyzed two synthetic EEG data sets, one generated such that each profile had one specific delay per trial, i.e., based on the shiftCP model, the other generated from the proposed convCP model. Secondly, we analyzed an event related EEG dataset based on a visual stimuli paradigm as well as a visual fMRI block design paradigm. To capture a strictly event related component in the data we constrained the last of the convCP components to be instantaneous. Then, the remaining convolutive components should model consistent confounding effects not phase locked to the event. For comparison we included a regular CP analysis as well as the shiftCP model proposed in [16], [24]. The convergence criterion for the algorithm was set as termination when the relative change of the log-posterior was less than 10^{-6} or when the algorithm had run for 1000 iterations. Since the optimization problems are non-convex the decompositions with highest log-posterior of 5 runs were chosen.

A. Simulated EEG data

The simulated data contain the five components[1] given to the left of figure 2. The data were sampled at 512 Hz and generated in 64 electrodes through 300 trials, white noise was added resulting in a SNR$= -10dB$. The components of the simulated data can be seen in figure 2 . The components are convolved with a convolutive filter with a total of 250 lags generated such that only 5% of the filter coefficients were non-zero (uniformly distributed).

From figure 2 it can be seen that the inadequate instantaneous CP model find a degenerate solution mainly capturing the activity of the true components 3 and 4. The convCP model is able to correctly identify the underlying components while the shiftCP model can only allocate one specific delay, thus, split the true component 4 into three components representing three different delays of the component time series rather than capturing true component 2 and 3. The CP model has degenerated, thus, allocates four components to account for the true component 4 and is unable to account for the remaining three convolutive components.

B. EEG: A visual object recognition study

This analysis is based on a visual object identification task [2]. The data were down sampled to 512 Hz and referenced to digitally linked earlobes. No trials were rejected – instead the data were high-pass filtered $>$ 3Hz to remove the most heavily confounding drift and slow wave effects. 50 Hz electronic noise was projected out using a multiple linear regression filter in intervals of 2 seconds. The data were cut into trials

[1]The first component corresponds to a mixture of a 40 Hz gamma burst with a 6 Hz theta activity located at occipital electrodes. The second component is a longer lasting 20 Hz beta burst most prominent at the central electrodes. The third component is a 5 Hz frontal theta burst. The fourth component is a frontal-occipital mixture of 12 and 13 Hz alpha activity overlayed with a 3 Hz delta activity. Finally, the 5th event related component is a mixture of 8.8 and 11 Hz alpha bursts most prominent in the occipital electrodes.

[2]The data contain a 64 channel recording of a healthy male subject based on the visual stimulus paradigm [14]. The paradigm consists of two types of black and white drawings: (1) objects (Ob), which are easily recognizable everyday type of objects like a chair, a number or a pipe, and (2) non-objects (Nob), which are random re-arrangements of the Ob drawings. Each stimulus category included 313 events and an object was presented up to three times. For details on the data set, see also [24].

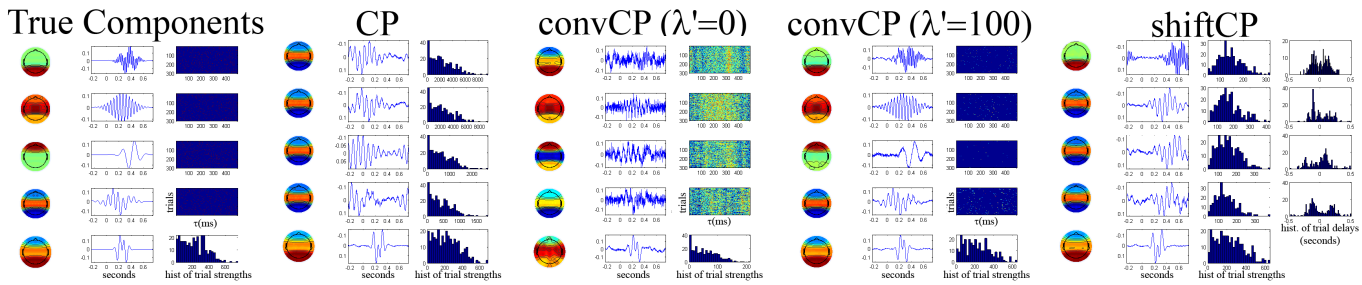

Fig. 2. A 5 component CP, convCP and shiftCP model of the data set generated by the convCP components to the far left. Noise was added to the data yielding a SNR$= -10dB$.

-500 to 1500 ms forming the data array $I = 64$ *channel* $\times$ $J = 1024$ *time-points* $\times$ $K = 313$ *trials*. The convCP analysis ($T = 250$ lags) as well as the corresponding instantaneous CP and shiftCP analysis of the data are given in figure 3. From the figure it can be seen that the instantaneous CP model has found a degenerate solution in which the activity of the eye-blink has been captured in the four first components (notice due to the high pass filtering only the higher frequency parts of the eye blink are modeled). The convCP and shiftCP model on the other hand does not form degenerate solutions. Instead 5 components are found. The first pertains to eye blinks. The second and third components pertain to occipital alpha activity while the fourth model central-frontal alpha activity. The fifth component particular for the convCP and shiftCP models seem to have well captured the EP both in terms of time profile and spatial distribution. Notice, how pruning in the convolutive filters (i.e. $\lambda' = 100$) makes the convCP resemble the shiftCP model, i.e. resulting in components pertaining to specific delays.

C. fMRI: BOLD response to visual stimulus

This analysis is based on a rapid acquisition visual stimulus fMRI data set designed to test models of the hemodynamic response in visual cortex[3]. The data set consists of 10 runs and has been preprocessed (including de-trending) as described in [10]. Thus, the size of the data was $I = 3891$ *voxels* $\times$ $J = 120$ *time-points* $\times$ $K = 10$ *trials*. To cover the stimulation period we set $T = 40$ samples.

In figure 4 a CP, convCP and shiftCP and convCP analysis of the data is given. All methods are able to extract the visual activity, however, in the convCP model the instantaneous component is more localized in the visual regions and has a less noisy temporal profile than the shiftCP model. Contrary to the CP and shiftCP decompositions no crosstalk between components is found in the convCP. When pruning the filter coefficients ($\lambda' = 100$) the convCP model does not reduce to the shiftCP model. Instead the convolutive components model

[3]A single slice was acquired at a sampling rate of 3 Hz in a para-axial orientation parallel to the calcarine sulcus using T2*-weighted EPI (1.5T). Each trial consists of 10 seconds (30 scans) of fixation, 10 seconds stimulation and 20 seconds of post-stimulus fixation. The visual stimulus was delivered as an annular full-field checkerboard reversing at 8 Hz. This is a strong stimulus for the primary visual areas that creates transient color and motion artifacts with activation over an extended set of visual areas.

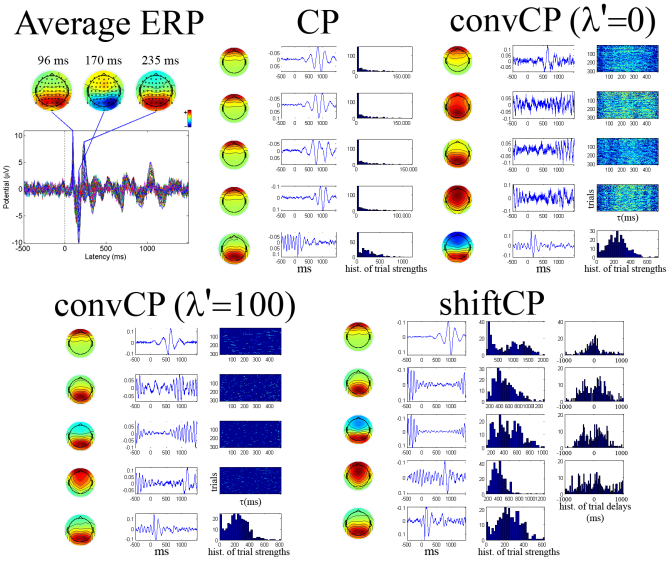

Fig. 3. **Tri-linear analysis of the EEG-data set**: (Top left panel) Average evoked potential (EP) for the 64 channels as well as spatial distribution at maximum amplitudes located at the latencies 96, 170 and 235 ms of the visually evoked $P100 - N200 - P300$ complex using the EEGLAB software [6]. (Top middle, right and bottom panel) A 5 component CP, convCP and shiftCP model of the EEG data set.

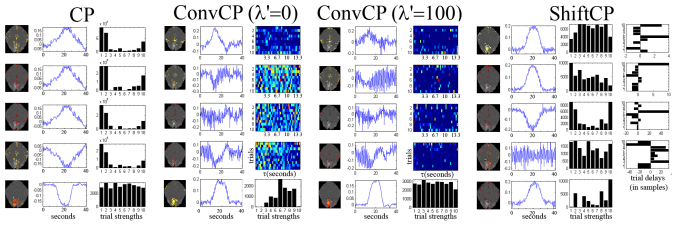

Fig. 4. **Tri-linear analyses of the fMRI data**: A 5 component CP, convCP and shiftCP analysis of the fMRI data set. The spatial profiles are thresholded to present the 5% of pixels with highest activity. Given are also the component time series as well as trial strengths, convolutive filters and delays for the various models respectively.

cardiac cycle effects prominent at frequencies around 0.8-1.2 Hz as well as low frequency drift. Thus, the convCP model more adequately model the data than the CP and shiftCP models.

IV. Discussion

. In both simulated and real EEG data as well as the fMRI data we find that the instantaneous tri-linear decomposition resulted in degenerated components. Thus, delay modelling is important for multi-linear decompositions of trial based neuroimaging data. While the components of the convCP model resembles the components of the the shiftCP model when an adequate degree of sparseness is imposed on the filter coefficients, the convCP model is in general more flexible than the shiftCP model. We see this both for the simulated convCP data as well as the fMRI data where the shiftCP model contrary to the convCP split components of similar time profiles into several components making the model loose information about other types of prominent activities in the data. Thus, for the fMRI data the convCP is able to both model the visual activity well while at the same time finding components related to the cardiac cycle.

Neuroimaging operates under great uncertainty. Not only are neuroimaging signals obtained under poor SNRs, the underlying responses to repeated stimuli are only partly reproducible, see e.g., [25]. In this work we have proposed a convolutive method convCP that exploit the tri-linear structure of the data and thereby identify consistent activities across the trials with the flexibility for delay and shape variation.

The convCP model generalizes directly to data of more modalities than three that naturally arise for instance when including modes such as subjects, conditions or runs [1] and also to convolutions across additional modes. While the focus in this presentation was on EEG and fMRI, the model should be readily applicable to other types of neuroimaging data such as magneto-encephalography and positron emission tomography. In this work we have focused on algorithmic issues related to MAP estimation and comparisons between different tri-linear models. Within the probabilistic framework we could apply standard approximate Bayesian methods for tuning hyperparameters (σ^2, λ) and to perform model selection [23], the matlab code for convCP is available for download from www.mortenmorup.dk.

References

[1] A. H. Andersen and W. S. Rayens. Structure-seeking multilinear methods for the analysis of fmri data. *NeuroImage*, 22:728–739, 2004.

[2] C.F. Beckmann and S.M. Smith. Tensorial extensions of independent component analysis for multisubject fmri analysis. *NeuroImage 25*, pages 294–311, 2005.

[3] A. J. Bell and T. J. Sejnowski. An information maximization approach to blind source separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[4] T. Blumensath and M. Davies. On shift-invariant sparse coding. *International Conference on Independent Component Analysis and Blind Source Separation*, 26:1205–1212, 2004.

[5] J. D. Carroll and J. J. Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35:283–319, 1970.

[6] A. Delorme and S.Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134:9–21, 2004.

[7] V. de Silva, and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.*, to appear, 2008.

[8] E. Donchin and E. Heffley. Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary perspectives in event-related brain potential research*, pages pp. 555–572, 1978.

[9] A. S. Field and D. Graupe. Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topogr.*, 3(4):407–423, 1991.

[10] C. Goutte, L. K. Hansen, M. G. Liptrot, and E. Rostrup. Feature-space clustering for fmri meta-analysis. *Human Brain Mapping*, 13:165–183, 2001.

[11] R. A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[12] R. A. Harshman and M. E. Lundy. Data preprocessing and the extended parafac model. *Research Methods for Multimode Data Analysis, Praeger, New York,*, pages 216–281, 1984.

[13] R. A. Harshman. Determination and proof of minimum uniqueness conditions for parafac1. *UCLA Working Papers in Phonetics*, (22):111–117, 1972.

[14] C.S. Herrmann, D. Lenz, S. Junge, N.A. Busch, and B. Maess. Memory-matches evoke human gamma-responses. *BMC Neuroscience*, 5(13), 2004.

[15] S. Hong and R. A. Harshman. Shifted factor analysis part iii: N-way generalization and application. *Journal of Chemometrics*, 17:389–399, 2003.

[16] K.H. Knuth, A.S. Shah, W.A. Truccolo, M. Ding, S.L. Bressler, and C.E. Schroeder. Differentially variable component analysis: Identifying multiple evoked components using trial-to-trial variability. *J. Neurophysiol.*, 95:3257–3276, 2006.

[17] J.B Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.*, 18:95–138, 1977.

[18] M. S. Lewicki and T. J. Sejnowski. Coding time-varying signals using sparse shift-invariant representations. *Adv. Neural Inform. Process. Systems (NIPS'99)*, 11:730–736, 1999.

[19] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. *Adv. in Neural Inform. Process. Systems (NIPS'96)*, 8:145–151, 1996.

[20] M. J. McKeown, T. P. Jung, S. Makeig, G. Brown, S. S. Kindermann, T. W. Lee, and T. J. Sejnowski. Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc. Natl. Acad. Sci. USA*, 95(3):803–810, Feb 1998.

[21] J. Möcks. Topographic components model for event-related potentials and some biophysical considerations. *IEEE Trans. Biomed. Eng.*, 35:482–484, 1988.

[22] M. Mørup. Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):24–40, 2011

[23] M. Mørup, L. K. Hansen. Automatic Relevance Determination for Multiway Models. *Journal of Chemometrics, Special Issue: In Honor of Professor Richard A. Harshman*, 23(7-8):352-363, 2009

[24] M Mørup, L.K. Hansen, S.M. Arnfred, L.-K. Lim, and K.H. Madsen. Shift invariant multilinear decomposition of neuroimaging data. *NeuroImage*, 42(4): 1439–1450, 2008.

[25] T. B. Penney, A. Mecklinger, and D. Nessler. Repetition related erp effects in a visual object target detection task. *Cognitive Brain Research*, 10(3):239–250, 2001.

[26] P. Smaragdis. Non-Negative Matrix Factor Deconvolution; Extraction of Multiple Sound Sourses from Monophonic Inputs. In *International Symposium on Independent Component Analysis and Blind Source Separation (ICA)*, 3195, 494–499,2004.

[27] M. P. Syskind, J. Larsen, U. Kjems, and L. C. Parra. A survey of convolutive blind source separation methods. *Springer Handbook on Speech Processing and Speech Communication*, 2007.

[28] K. Torkkola. Blind separation of delayed sources based on information maximization. *Acoustics, Speech, and Signal Processing. ICASSP-96*, 6:3509–3512, 1996.

[29] A. Yeredor. Time-delay estimation in mixtures. *ICASSP*, 5:237–240, 2003.